



(12)发明专利申请

(10)申请公布号 CN 106952190 A

(43)申请公布日 2017.07.14

(21)申请号 201710174076.X

(22)申请日 2017.03.22

(71)申请人 国信优易数据有限公司

地址 100070 北京市丰台区南四环西路188
号总部公馆(ABP)B座9楼

(72)发明人 金士隆 段立新 夏耘海

(74)专利代理机构 北京青松知识产权代理事务
所(特殊普通合伙) 11384

代理人 郑青松

(51)Int.Cl.

G06Q 50/16(2012.01)

权利要求书2页 说明书7页 附图1页

(54)发明名称

虚假房源录入行为识别与预警系统

(57)摘要

本发明公开了一种虚假房源录入行为识别与预警系统,包括:数据获取模块、行为识别模块、预警模型构建模块、目标预警模型确定模块和预警管理模块。本发明提供的虚假房源录入行为识别与预警系统能够对房产经纪人是否录入了虚假房源进行识别,并能对该行为进行预警,在预警值达到一定等级时,对该经纪人采取惩罚措施,以避免给用户带来不必要的麻烦和对同行业竞争对手造成严重的影响,防止影响公司名誉和形象,从而避免不必要的经济损失,规范房产交易市场。

数据获取模块1

行为识别模块2

预警模型构建模块3

目标预警模型确定模块4

预警管理模块5

1. 一种虚假房源录入行为识别与预警系统,其特征在于,包括:

数据获取模块,获取以预定时间段为周期记录的房源录入总量I和销售房源的如下信息:举报有效量Y、被他人改为无效的量A、被他人改为暂不出售的量B、被本人改为无效的量C、被本人改为暂不出售的量D,其中本人是指与获取的房源数据所对应的房产经纪人,他人是指除该房产经纪人之外的人;

行为识别模块,将获取的以预定时间段为周期记录的房源数据与预设条件进行对比,判断与获取的房源数据所对应的房产经纪人是否存在虚假房源录入行为,在判断为存在虚假房源录入行为的情况下,将该房产经纪人的识别标签标注为指示有虚假房源录入行为值1,否则将识别标签标注为值0;

预警模型构建模块,选取以预定时间段为周期记录的举报有效量、被他人改为无效的量、被他人改为暂不出售的量、被本人改为无效的量、被本人改为暂不出售的量作为特征向量,建立特征向量和虚假房源录入行为概率之间的逻辑回归模型,得到的预警模型如下式(1)所示:

$$\text{Logit}(P) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \theta_5 x_5 \quad (1)$$

其中,Logit(P)为预警值, x_1, x_2, x_3, x_4, x_5 分别为以预定时间段为周期记录的举报有效量、被他人改为无效的量、被他人改为暂不出售的量、被本人改为无效的量、被本人改为暂不出售的量的数值; θ_0 为截距, $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ 为各特征向量与虚假房源录入行为之间的关联度系数;

目标预警模型确定模块,基于获取的各特征向量的数值和所对应的识别标签值,对上述预警模型进行逻辑回归计算以得到各特征向量的回归值,从而得到进行预测的目标预警模型:

预警管理模块,在预定周期内,利用数据获取模块获取的房源数据和确定的目标预警模型来确定所获取的房源数据所对应的房产经纪人的虚假房源录入行为的预警值。

2. 根据权利要求1所述的虚假房源录入行为识别与预警系统,其特征在于,当以预定时间段为周期记录的房源数据满足下列预设条件的任一条件时,判断房产经纪人存在虚假房源录入行为:

$$(1) A \geq 5;$$

$$(2) B \geq 30;$$

$$(3) \frac{(A+C)}{I} \geq 10\% ;$$

$$(4) D \geq 10;$$

$$(5) Y \geq 5 \text{ 且 } \frac{Y}{I} \geq 10\% .$$

3. 根据权利要求2所述的虚假房源录入行为识别与预警系统,其特征在于,所述预定时间段为7天。

4. 根据权利要求3所述的虚假房源录入行为识别与预警系统,其特征在于,通过极大似然估计方法来确定构建的预警模型的各特征向量的回归值,得到的目标预警模型如下式(2)所示:

$$\text{Logit}(P) = -4.463 + 1.261x_1 + 0.16x_2 - 0.157x_3 + 0.162x_4 + 0.955x_5 \quad (2).$$

5. 根据权利要求1所述的虚假房源录入行为识别与预警系统，其特征在于，当预警值 Logit (P) 大于0时，将该房产经纪人的预警标签标注为预警状态；当预警值Logit (P) 小于0时，将该房产经纪人的预警标签标注为正常状态，如果该房产经纪人的预警次数或者预警等级超过一定阈值，则对该房产经纪人采取惩罚措施。

6. 根据权利要求5所述的虚假房源录入行为识别与预警系统，其特征在于，按照预警值 Logit (P) 的分布情况，按从低到高的顺序设置10个预警等级P1,P2,P3,P4,P5,P6,P7,P8,P9,P10，如果房产经纪人的预警等级连续3个月都为P10等级，则对该房产经纪人采取惩罚措施。

7. 根据权利要求6所述的虚假房源录入行为识别与预警系统，其特征在于，

当预警值Logit (P) 位于区间 (0, 200] 时，将预警等级设置为P1等级；

当预警值Logit (P) 位于区间 (200, 400] 时，将预警等级设置为P2等级；

当预警值Logit (P) 位于区间 (400, 600] 时，将预警等级设置为P3等级；

当预警值Logit (P) 位于区间 (600, 800] 时，将预警等级设置为P4等级；

当预警值Logit (P) 位于区间 (800, 1000] 时，将预警等级设置为P5等级；

当预警值Logit (P) 位于区间 (1000, 2000] 时，将预警等级设置为P6等级；

当预警值Logit (P) 位于区间 (2000, 4000] 时，将预警等级设置为P7等级；

当预警值Logit (P) 位于区间 (4000, 6000] 时，将预警等级设置为P8等级；

当预警值Logit (P) 位于区间 (6000, 9000] 时，将预警等级设置为P9等级；

当预警值Logit (P) 大于9000时，将预警等级设置为P10等级。

虚假房源录入行为识别与预警系统

技术领域

[0001] 本发明涉及一种虚假房源录入行为识别与预警系统,用于对虚假房源及录入行为进行识别。

背景技术

[0002] 目前的房产信息大都是通过房产中介公司的房源展示系统来进行,该系统向房地产经纪人提供一个输入端口,房产经纪人可以将房产相关信息录入到这个系统中供用户选择。用户通过浏览这些房产信息来决定是否购买,同时户主也可以委托房产经纪人将自己的房产信息放到平台上进行售卖,我们将这样一个系统称之为房源展示系统。

[0003] 这个房源展示系统由某个特定的房产中介公司实际运营,房产中介公司对房地产经纪人进行考核,考核的指标中包括房源录入次数、核实施房源信息真实性次数、售出房源套数等指标;因为涉及到业绩的考核,所以一些房产经纪人会录入一些虚假编造的房源来满足考核的要求,而这些虚假的房源信息若没有经过仔细的勘察检验会最终被展示到房源信息页面中供用户浏览,为用户购置房产带来了不必要的麻烦,同时也因表面上的房源信息丰富而对同行竞争对手造成严重的影响,更因为房价信息的胡乱编造而对整个社会的房价带来十分恶劣的影响。

[0004] 基于以上原因,构造监督式学习算法通过识别虚假房源信息对房产经纪人录入虚假房源的行为进行预警,从而降低由此带来的影响。

发明内容

[0005] 针对上述技术问题,本发明提供一种能够识别虚假房源录入行为并对该行为进行预警的虚假房源录入行为识别与预警系统。

[0006] 本发明采用的技术方案为:

[0007] 本发明的实施例提供一种虚假房源录入行为识别与预警系统,包括:

[0008] 数据获取模块,获取以预定时间段为周期记录的房源录入总量I和销售房源的如下信息:举报有效量Y、被他人改为无效的量A、被他人改为暂不出售的量B、被本人改为无效的量C、被本人改为暂不出售的量D,其中本人是指与获取的房源数据所对应的房产经纪人,他人是指除该房产经纪人之外的人;

[0009] 行为识别模块,将获取的以预定时间段为周期记录的房源数据与预设条件进行对比,判断与获取的房源数据所对应的房产经纪人是否存在虚假房源录入行为,在判断为存在虚假房源录入行为的情况下,将该房产经纪人的识别标签标注为指示有虚假房源录入行为值1,否则将识别标签标注为值0;

[0010] 预警模型构建模块,选取以预定时间段为周期记录的举报有效量、被他人改为无效的量、被他人改为暂不出售的量、被本人改为无效的量、被本人改为暂不出售的量作为特征向量,建立特征向量和虚假房源录入行为概率之间的逻辑回归模型,得到的预警模型如下式(1)所示:

[0011] $\text{Logit}(P) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_4 + \theta_5 X_5 \quad (1)$

[0012] 其中, $\text{Logit}(P)$ 为预警值, X_1, X_2, X_3, X_4, X_5 分别为以预定时间段为周期记录的举报有效量、被他人改为无效的量、被他人改为暂不出售的量、被本人改为无效的量、被本人改为暂不出售的量的数值; θ_0 为截距, $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ 为各特征向量与虚假房源录入行为之间的关联度系数;

[0013] 目标预警模型确定模块, 基于获取的各特征向量的数值和所对应的识别标签值, 对上述预警模型进行逻辑回归计算以得到各特征向量的回归值, 从而得到进行预测的目标预警模型:

[0014] 预警管理模块, 在预定周期内, 利用数据获取模块获取的房源数据和确定的目标预警模型来确定所获取的房源数据所对应的房产经纪人的虚假房源录入行为的预警值。

[0015] 可选地, 当以预定时间段为周期记录的房源数据满足下列预设条件的任一条件时, 判断房产经纪人存在虚假房源录入行为:

[0016] (1) $A \geq 5$;

[0017] (2) $B \geq 30$;

[0018] (3) $\frac{(A+C)}{I} \geq 10\%$;

[0019] (4) $D \geq 10$;

[0020] (5) $Y \geq 5$ 且 $\frac{Y}{I} \geq 10\%$ 。

[0021] 可选地, 所述预定时间段为7天。

[0022] 可选地, 通过极大似然估计方法来确定构建的预警模型的各特征向量的回归值, 得到的目标预警模型如下式(2)所示:

[0023] $\text{Logit}(P) = -4.463 + 1.261X_1 + 0.16X_2 - 0.157X_3 + 0.162X_4 + 0.955X_5 \quad (2)$ 。

[0024] 可选地, 当预警值 $\text{Logit}(P)$ 大于 0 时, 将该房产经纪人的预警标签标注为预警状态; 当预警值 $\text{Logit}(P)$ 小于 0 时, 将该房产经纪人的预警标签标注为正常状态, 如果该房产经纪人的预警次数或者预警等级超过一定阈值, 则对该房产经纪人采取惩罚措施

[0025] 可选地, 按照预警值 $\text{Logit}(P)$ 的分布情况, 按从低到高的顺序设置 10 个预警等级 P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, 如果房产经纪人的预警等级连续 3 个月都为 P10 等级, 则对该房产经纪人采取惩罚措施。

[0026] 可选地, 当预警值 $\text{Logit}(P)$ 位于区间 $(0, 200]$ 时, 将预警等级设置为 P1 等级; 当预警值 $\text{Logit}(P)$ 位于区间 $(200, 400]$ 时, 将预警等级设置为 P2 等级; 当预警值 $\text{Logit}(P)$ 位于区间 $(400, 600]$ 时, 将预警等级设置为 P3 等级; 当预警值 $\text{Logit}(P)$ 位于区间 $(600, 800]$ 时, 将预警等级设置为 P4 等级; 当预警值 $\text{Logit}(P)$ 位于区间 $(800, 1000]$ 时, 将预警等级设置为 P5 等级; 当预警值 $\text{Logit}(P)$ 位于区间 $(1000, 2000]$ 时, 将预警等级设置为 P6 等级; 当预警值 $\text{Logit}(P)$ 位于区间 $(2000, 4000]$ 时, 将预警等级设置为 P7 等级; 当预警值 $\text{Logit}(P)$ 位于区间 $(4000, 6000]$ 时, 将预警等级设置为 P8 等级; 当预警值 $\text{Logit}(P)$ 位于区间 $(6000, 9000]$ 时, 将预警等级设置为 P9 等级; 当预警值 $\text{Logit}(P)$ 大于 9000 时, 将预警等级设置为 P10 等级。

[0027] 本发明提供的虚假房源录入行为识别与预警系统能够对房产经纪人是否录入了虚假房源进行识别, 并能对该行为进行预警, 在预警值达到一定等级时, 对该经纪人采取惩

罚措施,以避免给用户带来不必要的麻烦和对同行业竞争对手造成严重的影响,防止影响公司名誉和形象,从而避免不必要的经济损失,规范房产交易市场。

附图说明

[0028] 图1为本发明的虚假房源录入行为识别与预警系统的结构示意图。

具体实施方式

[0029] 为使本发明要解决的技术问题、技术方案和优点更加清楚,下面将结合附图及具体实施例进行详细描述。

[0030] 图1为本发明的虚假房源录入行为识别与预警系统的结构示意图。

[0031] 如图1所示,本发明提供的一种虚假房源录入行为识别与预警系统,用于根据销售房屋的公司的数据库所采集的数据来构建用于对房产经纪人录入虚假房源的行为进行监测的预警模型,从而可对房产经纪人的行为进行规范化管理,包括数据获取模块1、行为识别模块2、预警模型构建模块3和目标预警模型确定模块。以下分别对这些模块进行详细介绍。

[0032] 【数据获取模块】

[0033] 数据获取模块1用于获取房源录入总量I和销售房源的如下信息:获取以预定时间段为周期记录的房源录入总量I和销售房源的如下信息:举报有效量Y、被他人改为无效的量A、被他人改为暂不出售的量B、被本人改为无效的量C、被本人改为暂不出售的量D,其中本人是指与获取的房源数据所对应的房产经纪人,他人是指除该房产经纪人之外的人。由于目前房产中介公司的房源状态在后台更改且被审核的周期最长为7天,所以在发明的一示意性实施例中,可将预定时间段设置为7天,这样可在后续预警模型的构建中建立以7为时长的变量,但并不局限于此。

[0034] 在实际操作中,房产经纪人通过房源展示系统中的电话系统和出售人进行沟通,系统存下录音,通过语音识别、自然语言处理等方法将该通话进行标签标注,如无人接听、占线、暂不出售、无效等标签,然后系统存入数据库。后台数据库中存有房源状态被修改的时间日期以及修改人,通过简单的数据逻辑计算生成新的字段并保存新数据库表:房源录入总量、举报有效量、被他人改为无效的量、被他人改为暂不出售的量、被本人改为无效的量、被本人改为暂不出售的量、虚假录入状态、预警值。

[0035] 这样,数据获取模块1可从后台数据库获取某段时间的数据进行行为识别和用于预警模型构建,并且可在构建预警模型之后按照预定周期,例如月末采集本月的房源数据来对房产经纪人的房源录入行为进行管理。

[0036] 【行为识别模块】

[0037] 行为识别模块2用于将获取的以预定时间段为周期记录的房源数据与预设条件进行对比,判断与获取的房源数据所对应的房产经纪人是否存在虚假房源录入行为,在判断为存在虚假房源录入行为的情况下,将该房产经纪人的识别标签标注为指示有虚假房源录入行为值1,否则将识别标签标注为值0。

[0038] 具体地,当以预定时间段为周期记录的房源数据满足下列预设条件的任一条件时,即只要下述条件有一条得到满足,则判断房产经纪人存在虚假房源录入行为:

[0039] (1) 被他人改为无效的量 ≥ 5 , 即 $A \geq 5$;

[0040] (2) 被他人改为暂不出售的量 ≥ 30 , 即 $B \geq 30$;

[0041] (3) (被他人改为无效的量+被本人改为无效的量)/房源录入总量 $\geq 10\%$, 即

$$\frac{(A+C)}{I} \geq 10\% ;$$

[0042] (4) 被本人改为暂不出售的量 ≥ 10 , 即 $D \geq 10$;

[0043] (5) 举报有效量 ≥ 5 且举报有效量/房源录入总量 $\geq 10\%$, 即 $Y \geq 5$ 且 $\frac{Y}{I} \geq 10\%$ 。

[0044] 在实际操作中, 当根据获取的数据确认某房产经纪人存在虚假房源录入行为时, 则在后台数据库中将该房产经纪人的ID进行标注。

[0045] 【预警模型构建模块】

[0046] 预警模型构建模块3用于选取数据获取模块1所获取的举报有效量Y、被他人改为无效的量A、被他人改为暂不出售的量B、被本人改为无效的量C、被本人改为暂不出售的量D作为特征向量, 建立特征向量和虚假房源录入行为概率之间的逻辑回归模型。

[0047] 具体地, 将房源数据的特征集合x作为逻辑回归模型的独立变量向量, $x = \{x_1, x_2, x_3, x_4, x_5\}$, 虚假房源录入行为y作为逻辑回归模型的因变量, 取值为0或1, 这样, 引入将回归结果y二分化的激励函数(sigmoid函数) $g(x)$, $g(x) = \frac{1}{(1 + e^{-y})} = \frac{1}{(1 + e^{-h(x)})}$, $h(x) = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$ 构造发生虚假房源录入行为的概率函数为:

$$[0048] P(y=1|x; \theta) = g(x) = \frac{1}{(1 + e^{-h(x)})}$$

[0049] 不发生虚假房源录入行为的概率函数为:

$$[0050] P(y=0|x; \theta) = 1 - g(x) = \frac{1}{1 + e^{h(x)}}$$

[0051] 可得事件的发生比:

$$[0052] \frac{P(y=1|x; \theta)}{P(y=0|x; \theta)} = e^{h(x)}$$

[0053] 又因为, $h(x) = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$, 所以对事件的发生比取对数, 得到线性函数, 即虚假房源录入行为的预警模型如下式(1)所示:

$$[0054] \text{Logit}(P) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \theta_5 x_5 \quad (1)$$

[0055] 其中, Logit(P)为虚假房源录入行为概率优势比的对数值, 又称为预警值, 预警值越大说明与获取的房源数据所对应的房产经纪人的录入虚假房源的行为次数越多, 行为越恶劣; x_1, x_2, x_3, x_4, x_5 分别为预定时间段内获取的举报有效量、被他人改为无效的量、被他人改为暂不出售的量、被本人改为无效的量、被本人改为暂不出售的量的数值, y为识别标签值, 取值为0时, 表示不存在虚假房源录入行为, 取值为1时, 表示存在虚假房源录入行为; θ_0 为截距, $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ 为各特征向量与虚假房源录入行为之间的关联度系数。

[0056] 【目标预警模型确定模块】

[0057] 目标预警模型确定模块4基于获取的各特征向量的数值和所对应的识别标签值, 对上述预警模型进行逻辑回归计算以得到各特征向量的回归值, 从而得到对虚假房源录入

行为进行预测的目标预警模型。

[0058] 具体地,本发明利用极大似然估计方法来确定构建的预警模型的各特征向量的回归值。基于上述公式(1)构建的似然函数如下述公式(2)所示:

$$[0059] L(\theta) = \prod_{i=1}^m p(y_i | x_i; \theta) = \prod_{i=1}^m g(x_i)^{y_i} (1-g(x_i))^{1-y_i} \quad (2)$$

[0060] 对上述公式(3)所示的极大似然函数两端求导得到其对数似然函数,如下式(4)所示:

$$[0061] l(\theta) = \sum y_i * \log(g(x_i)) + (1-y_i) * \log(1-g(x_i)) \quad (3)$$

[0062] 通过对上述公式(3)所示的对数似然函数的参数进行求解,可得到预警模型的模型参数,即回归值,从而得到最终的目标预测模型。

[0063] 本发明采用梯度下降法来对最大似然函数的参数进行求解,包括如下步骤:

[0064] (1)选择梯度方向 $\nabla J(\theta)$;

[0065] (2)选择下降步长 α ,对参数进行更新: $\theta^i = \theta^{i-1} - \alpha * \nabla J(\theta^{i-1})$

[0066] (3)重复以上步骤直至满足停止条件。

[0067] 其中, $\nabla J(\theta) = -\frac{1}{n} \sum_i (y_i - y_i^*) * x_i + \lambda \theta$, x_i 是投影到 $[0, 1]$ 之间的任意变量, y_i 是第 i 个样本的真实值, y_i^* 是该样本的预测值, λ 是防止模型过拟合的结构化系数。

[0068] 在实际操作中,可通过计算机软件如SAS软件来运用获取的举报有效量、被他人改为无效的量、被他人改为暂不出售的量、被本人改为无效的量、被本人改为暂不出售的量的数值以及对应的标签值来进行回归求得各特征向量的权重 $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ 。

[0069] 【预警管理模块】

[0070] 预警管理模块5用于在预定周期内,利用数据获取模块获取的房源数据和得到的目标预警模型来确定所获取的房源数据所对应的房产经纪人的虚假房源录入行为的预警值,当预警值Logit(P)大于0时,将该房产经纪人的预警标签标注为预警状态;当预警值Logit(P)小于0时,将该房产经纪人的预警标签标注为正常状态,如果该房产经纪人的预警次数或者预警等级超过一定阈值,则对该房产经纪人采取惩罚措施。例如,以一个月为周期,利用采集的房源数据和上述目标预警模型可得到与房源数据相对应的房产经纪人的预警值,当预警值Logit(P)大于0时,在后台数据库中将该房产经纪人的ID标注为预警状态,当预警值Logit(P)小于0时,标注为正常状态,并按照预警值Logit(P)的分布情况,按从低到高的顺序设置10个预警等级P1,P2,P3,P4,P5,P6,P7,P8,P9,P10。如果根据某段时间的数据计算的预警值越大说明此经纪人录入虚假房源的行为次数越多,行为越恶劣。如果房产经纪人的预警次数或者预警等级超过一定阈值时,则对该房产经纪人采取惩罚措施,例如,如果一个经纪人连续3个月的预警等级都为P10,则需要对其进行严重惩罚,具体的处罚措施可根据实际情况来确定,例如可开除该经纪人。具体地:

[0071] 当预警值Logit(P)位于区间(0,200]时,将预警等级设置为P1等级;

[0072] 当预警值Logit(P)位于区间(200,400]时,将预警等级设置为P2等级;

[0073] 当预警值Logit(P)位于区间(400,600]时,将预警等级设置为P3等级;

[0074] 当预警值Logit(P)位于区间(600,800]时,将预警等级设置为P4等级;

[0075] 当预警值Logit(P)位于区间(800,1000]时,将预警等级设置为P5等级;

- [0076] 当预警值Logit (P) 位于区间 (1000,2000] 时,将预警等级设置为P6等级;
 [0077] 当预警值Logit (P) 位于区间 (2000,4000] 时,将预警等级设置为P7等级;
 [0078] 当预警值Logit (P) 位于区间 (4000,6000] 时,将预警等级设置为P8等级;
 [0079] 当预警值Logit (P) 位于区间 (6000,9000] 时,将预警等级设置为P9等级;
 [0080] 当预警值Logit (P) 大于9000时,将预警等级设置为P10等级。

[0081] 【实施例】

[0082] 本发明以某大型房产中介公司的后台数据库所记录的数据为例,对本发明的虚假房源录入行为识别与预警系统进行说明。

[0083] 在本实施例的目标预警模型的确定中,使用该大型房产中介公司的2015年1月到2016年2月份所统计的北京、南京、上海、深圳、苏州、天津、武汉、长春、重庆等9个重点城市的48986条数据作为样本数据,如下表1所示。

[0084] 表1:房源数据记录表

[0085]

序号	月份	城市	房源ID	录入总量	虚假房源量	录入总量占比	虚假房源占比	虚假房源占比下限	虚假房源占比上限	预警状态	预警等级	预测值	误差
2015	12	南京	5385420	888	828	9%	9%	8%	10%	1	P6	796.27	
2015	12	深圳	3388559	725	238	3%	3%	2%	4%	1	P6	225.38	
2015	12	上海	2436544	38741	6	0.0%	0.0%	0.0%	0.0%	1	P6	333.48	
2015	12	上海	5385420	12293	106	0.8%	0.8%	0.5%	1.1%	1	P6	356.93	
2015	12	杭州	2388264	3885	182	4%	4%	3%	5%	1	P6	346.51	
2015	12	青岛	5315237	2288	182	8%	8%	7%	9%	1	P6	339.98	
2015	12	苏州	5322388	639	188	3%	3%	2%	4%	1	P6	222.52	
2015	12	南京	2388132	1113	178	3%	3%	2%	4%	1	P6	323.98	
2015	12	上海	2733923	613	178	3%	3%	2%	4%	1	P6	218.31	
2015	12	上海	2738473	16629	78	0.5%	0.5%	0.2%	1.1%	1	P6	218.38	
2015	12	上海	2719322	522	178	3%	3%	2%	4%	1	P6	212.33	
2015	12	南京	3388559	437	178	4%	4%	3%	5%	1	P6	238.71	
2015	12	南京	5385420	780	188	3%	3%	2%	4%	1	P6	208.67	
2015	12	南京	3388559	298	183	6%	6%	5%	7%	1	P6	208.36	
2015	12	杭州	2389582	988	191	2%	2%	1%	3%	1	P6	325.68	
2015	12	杭州	5385420	2288	188	8%	8%	7%	9%	1	P6	338.38	
2015	12	上海	2377657	2028	152	7%	7%	5%	9%	1	P6	338.55	
2015	12	上海	2377657	1928	188	10%	10%	7%	13%	1	P6	338.38	
2015	12	上海	2734150	12288	49	0.4%	0.4%	0.2%	0.8%	1	P6	186.76	
2015	12	上海	2734150	12288	131	1%	1%	0.5%	2%	1	P6	186.76	
2015	12	南京	5385420	678	148	2%	2%	1%	3%	1	P6	332.97	
2015	12	杭州	5383388	1529	148	1%	1%	0.5%	2%	1	P6	322.98	
2015	12	南京	2385422	374	148	4%	4%	2%	6%	1	P6	329.98	
2015	12	南京	5382426	622	148	3%	3%	2%	4%	1	P6	472.57	
2015	12	苏州	5383388	439	138	1%	1%	0.5%	2%	1	P6	376.52	
2015	12	南京	2385521	367	137	3%	3%	2%	4%	1	P6	338.83	
2015	12	南京	5315548	438	137	3%	3%	2%	4%	1	P6	338.38	
2015	12	武汉	5385526	173	138	3%	3%	2%	4%	1	P6	351.88	
2015	12	南京	3388559	348	138	3%	3%	2%	4%	1	P6	338.38	
2015	12	天津	5385526	698	133	2%	2%	1%	3%	1	P6	338.38	
2015	12	南京	3389742	528	138	3%	3%	2%	4%	1	P6	344.53	
2015	12	天津	2388273	648	138	2%	2%	1%	3%	1	P6	343.87	
2015	12	南京	5385526	888	138	3%	3%	2%	4%	1	P6	342.68	
2015	12	上海	2732221	138	316	3%	3%	2%	4%	1	P6	336.33	
2015	12	南京	5382426	498	313	2%	2%	1%	3%	1	P6	336.68	
2015	12	天津	2399557	865	316	3%	3%	2%	4%	1	P6	336.53	
2015	12	南京	5388976	888	312	3%	3%	2%	4%	1	P6	332.93	

[0086] 利用表1记录的数据,运用梯度下降法对构建的预测模型进行参数求解,迭代次数为11次,步长为1,得到的回归结果如下表2所示:

[0087] 表2:回归结果

[0088] 方程式中的变量

[0089]

	B	S.E.	Wald	自由度	显著性	Exp(B)
步骤 1 ^a 假设检验	1.261	.036	1326.435	1	.000	3.529
本人7天内改为暂不出售	.160	.044	13.325	1	.000	1.174
别人7天内改为暂不出售	-0.157	.042	13.863	1	.000	0.858
本人7天内改为无效	.162	.018	78.241	1	.000	1.176
别人7天内改为无效	.955	.016	3830.373	1	.000	2.599
常量	-4.463	.048	2849.782	1	.000	.012

[0090] a. 步骤1: [%] :,1:

[0091] 由此可得到进行预测的目标预警模型如下式(4)所示:

[0092] $\text{Logit}(P) = -4.463 + 1.261X_1 + 0.16X_2 - 0.157X_3 + 0.162X_4 + 0.955X_5 \quad (4)$

[0093] 此外,由表2可知,本实施例建立的目标预警模型的显著性检验效果非常好,从而说明根据本发明提供的虚假房源录入行为识别与预警系统所建立的目标预警模型非常合理。

[0094] 以上所述是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明所述原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也应视为本发明的保护范围。



图1